Tight Bound on Randomness for Violating the Clauser-Horne-Shimony-Holt Inequality

Yifeng Teng, Shenghao Yang, *Member, IEEE*, Siwei Wang and Mingfei Zhao

***Abstract*—Free will (or randomness) has been studied to achieve loophole-free Bell's inequality test and to provide device-independent quantum key distribution security proofs. The required randomness such that a local hidden variable model (LHVM) can violate the Clauser-Horne-Shimony-Holt (CHSH) inequality has been studied, but a tight bound has not been proved for a practical case that i) the device settings of the two parties in the Bell test are independent; and ii) the device settings of each party can be correlated or biased across different runs. Using some information theoretic techniques, we prove in this paper a tight bound on the required randomness for this case such that the CHSH inequality can be violated by certain LHVM. Our proof has a clear achievability and converse style. The achievability part is proved using type counting. To prove the converse part, we introduce a concept called profile for a set of binary sequences and study the properties of profiles. Our profile-based converse technique is also of independent interest.**

***Index Terms*—Bell's inequality test, CHSH inequality, randomness loophole, randomness bound**

I. Introduction

Bell's inequality test [1] provides an approach to verify the existence of physical phenomenon that cannot be explained by local hidden variable models (LHVMs). The Clauser-Horne-Shimony-Holt (CHSH) inequality [2] is the most often used inequality in Bell test experiments. Experimental demonstrations of the violation of CHSH inequalities have been conducted since 1982 [3] (see also Giustina et al.'s work [4] and the references therein). These Bell tests, however, suffer from an inherent loophole that the settings of the participated devices may not be chosen totally randomly, called the *randomness (free will)* loophole. A small amount of correction between the device settings makes it possible that a LHVM can reproduce predictions of quantum mechanics [5]–[9]. This loophole also weakens the Bell's inequality based security proofs of device-independent quantum key distribution [10]–[12] and randomness expansion [13]–[15].

One of the essential questions in the randomness loophole is the bound of randomness such that the correctness of Bell tests can (or cannot) be guaranteed [7]–[9], [16]–[20]. Using a min-entropy type randomness measure, the bound of randomness required in a CHSH inequality test can be formulated as an optimization problem, and various special cases have been solved [16], [18], [19]. One case that has not been completely resolved in the literature is that the two parties of the test have independent settings, but the setting of each party can be biased or correlated across different runs. In this paper, we study this case and obtain the asymptotic optimal value explicitly.

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61471215. This work was partially funded by a grant from the University Grants Committee of the Hong Kong Special Administrative Region (Project No. AoE/E-02/08).

Y. Teng is with the Department of Computer Sciences, University of Wisconsin-Madison, Madison, USA (e-mail: yifengt@cs.wisc.edu).

S. Yang is with the School of Science of Engineering, The Chinese University of Hong Kong, Shenzhen, P. R. China (e-mail: shyang@cuhk.edu.cn).

S. Wang is with the Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, P. R. China (e-mail: wangsw15@mails.tsinghua.edu.cn).

M. Zhao is with the School of Computer Science, McGill University, Montreal, Canada (e-mail: mingfei.zhao@mail.mcgill.ca).

A. Problem Formulation

Let n be a positive integer, and X, Y be two random variables over $\{0,1\}^n$ with a joint distribution p_{XY}. We may consider that X and Y are the device settings of the two parties in an n-run Bell test, respectively. The following randomness measure has been used in the literature:

$$P = \left(\max_{\mathbf{x},\mathbf{y}\in\{0,1\}^n} p_{XY}(\mathbf{x},\mathbf{y}) \right)^{1/n}.$$

When X and Y are independent and uniformly distributed, $P = 1/4$, which is the minimum value of P and corresponds to the case of complete randomness. When X and Y are deterministic, $P = 1$, which corresponds to the case of zero randomness. Note that P is related to the min-entropy:

$$H_\infty(X,Y) := -\log \max_{\mathbf{x},\mathbf{y}\in\{0,1\}^n} p_{XY}(\mathbf{x},\mathbf{y}) = -n \log P.$$

Regard the vectors $\mathbf{x} \in \{0,1\}^n$ as column vectors and denote by $\mathbf{x}^{\mathrm{T}}$ the transpose of the $\mathbf{x}$. The optimization problem of interest is

$$\begin{aligned} \min_{p_{XY}} \quad & P \\ \text{s.t.} \quad & \frac{1}{n} \sum_{\mathbf{x},\mathbf{y}} \mathbf{x}^{\mathrm{T}} \mathbf{y} p_{XY}(\mathbf{x},\mathbf{y}) \le \frac{4 - S_Q}{8}, \end{aligned} \tag{1}$$

where $S_Q = 2\sqrt{2}$ is a quantum constant. Readers may refer to [7], [16], [18] to see how this problem is obtained. Optimization (1) can be simplified to a linear programming [18]. When $n = 1$, the optimal value of (1) is $(S_Q + 4)/24 \approx 0.285$, which was shown by Hall [7] and Koh et al. [16]. When $n \to \infty$, Pope and Kay [18] showed that the optimal value of (1) converges to $3^{\frac{-S_Q-4}{8}} 2^{-h_{\mathrm{b}}\left(\frac{4-S_Q}{8}\right)} \approx 0.258$, where

$$h_{\mathrm{b}}(t) = -t \log_2 t - (1-t) \log_2 (1-t)$$

is the binary entropy function.

The case that X and Y are independent is of particular interest. Towards a loophole free Bell test, physicists have

designed experiments with independent device settings [21]. In quantum key distribution, the experimental devices of the two parties may be manufactured independently and separated spatially, reducing the potential correlation of the device settings generated by the adversary. For independent device settings, the corresponding optimization problem becomes

$$\begin{array}{ll}\min\limits_{p_X,p_Y} & P \\ \text{s.t.} & \dfrac{1}{n}\sum\limits_{\mathbf{x},\mathbf{y}} \mathbf{x}^{\mathrm{T}}\mathbf{y} p_{XY}(\mathbf{x},\mathbf{y}) \le \dfrac{4-S_Q}{8} \\ & p_{XY}(\mathbf{x},\mathbf{y}) = p_X(\mathbf{x})p_Y(\mathbf{y}).\end{array} \tag{2}$$

Note that the above problem is not derived by directly imposing the constraint $p_{XY}(\mathbf{x},\mathbf{y}) = p_X(\mathbf{x})p_Y(\mathbf{y})$ to (1). For the completeness, we briefly discuss how (2) is derived from the corresponding CHSH inequality test problem in Appendix I.

When $n = 1$, it was obtained by Koh et al. [16] that the optimal value of (2) is $S_Q/8 \approx 0.354$. Let P_Q be the limit of the optimal value of (2), when $n \to \infty$. The value of P_Q has the following interpretation. For any independent device settings with randomness less than P_Q, it is not possible to have a LHVM that violates CHSH inequality. But for any value $P > P_Q$, there exists a LHVM that violates CHSH inequality where the device settings are independent, but have randomness less than or equal to P. Therefore, we are motivated to study the value of P_Q for CHSH inequality test. Yuan, Cao and Ma [19] have shown numerically that $P_Q \lessapprox 0.264$.

B. Our Contribution

In this paper, we provide an exact characterization of P_Q, and hence close the unresolved case in Table I. Particularly, we show that

$$P_Q = 4^{-h_{\mathrm{b}}(\sqrt{c_Q})} = 0.26428\ldots,$$

where $c_Q = \frac{4-S_Q}{8} \approx 0.1464$. Our formula has a min-entropy interpretation: $-n\log_2 P_Q = 2nh_{\mathrm{b}}(\sqrt{c_Q})$, i.e., each bit in X and Y has an average min-entropy $h_{\mathrm{b}}(\sqrt{c_Q})$.

To prove achievability, we simplify (2) by introducing an extra constraint that both X and Y have the uniform distribution over $\mathcal{A}_{n,l}$, the set of sequences in $\{0,1\}^n$ with at most nl 1s, and obtain a new optimization problem

$$\begin{array}{ll}\min\limits_{l} & (1/|\mathcal{A}_{n,l}|)^{2/n} \\ \text{s.t.} & \dfrac{1}{n|\mathcal{A}_{n,l}|^2}\sum\limits_{\mathbf{x},\mathbf{y}\in\mathcal{A}_{n,l}} \mathbf{x}^{\mathrm{T}}\mathbf{y} \le \dfrac{4-S_Q}{8},\end{array} \tag{2$'$}$$

which is essentially the same problem studied in [19, Section IV-B]. The asymptotic optimal value of (2′) when $n \to \infty$, denoted by $\hat{P}_Q$, gives an upper bound on P_Q since (2′) is obtained by reducing the feasible region of (2). The numerical bound on $\hat{P}_Q$ in [19] can be made analytical, and it shows that $\hat{P}_Q \le 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$ and hence $P_Q \le 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$.

The major part of our paper is to show the converse that no distributions of X and Y with randomness less than $4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$ can be feasible for (2), i.e., $P_Q \ge 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$. Note that we cannot use (2′) as the starting point to prove the converse since the derivation of (2′) implies $\hat{P}_Q \ge P_Q$. It is possible to show that $\hat{P}_Q \ge 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$, but not $P_Q \ge 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$ by studying only (2′).

TABLE I
PREVIOUS RESULTS.

	correlated devices	independent devices
$n = 1$	$(S_Q+4)/24 \approx 0.285$	$S_Q/8 \approx 0.354$
$n \to \infty$	$3^{-\frac{S_Q+4}{8}} 2^{-h_{\mathrm{b}}\left(\frac{4-S_Q}{8}\right)} \approx 0.258$	$\lessapprox 0.264$

To prove converse, we introduce a concept called *profile* to characterize a set of binary sequences. We study some properties of profiles, based on which optimization (2) is simplified and the converse is proved. The technique of profile seems to be firstly used here and may of independent interest for other problems.

In the remainder of this paper, our techniques used to prove the main result are summarized in the next section, followed by the details in Section III. Some concluding remarks are given in Section IV.

II. Outline of the Proofs

As described in the previous section, we formulate an optimization problem as follows.

Problem 1. *For any given $c \in (0, 1/4]$ and every positive integer n, consider the following program*

$$\begin{array}{ll}\min\limits_{p_X,p_Y} & \left(\max\limits_{\mathbf{x}} p_X(\mathbf{x}) \max\limits_{\mathbf{y}} p_Y(\mathbf{y})\right)^{1/n}, \\ \textit{s.t.} & \dfrac{1}{n}\sum\limits_{\mathbf{x},\mathbf{y}\in\{0,1\}^n} p_X(\mathbf{x})p_Y(\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y} \le c,\end{array}$$

where p_X and p_Y are probability distributions over $\{0,1\}^n$. Let P_n be the optimal value of the above program. We are interested in the limit of the sequence $\{P_n\}$ when $n \to \infty$.

Specifically we will need the case that $c = c_Q$ for the physics problem of interests. Now we state the following theorem.

Theorem 1. *For Problem 1 with $c = c_Q$, $\lim\limits_{n\to\infty} P_n = 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$, where*

$$h_{\mathrm{b}}(t) = -t\log_2 t - (1-t)\log_2(1-t) \tag{3}$$

is the binary entropy function.

In the following of this section, we give an outline of the main techniques towards proving this theorem. We have the following bound for P_n.

Proposition 1. *For all sufficiently large n, $1/4 \le P_n < 1/2$.*

A. Simplified Problem

Let S_X and S_Y be the support of distributions p_X and p_Y, respectively. Problem 1 can be simplified if we only consider distributions that are uniform over support. Suppose that

$$\begin{aligned} p_X(\mathbf{x}) &= \frac{1}{|S_X|},\ \forall \mathbf{x} \in S_X, \\ p_Y(\mathbf{y}) &= \frac{1}{|S_Y|},\ \forall \mathbf{y} \in S_Y. \end{aligned}$$

Then we have

$$\left(\max_{\mathbf{x}} p_X(\mathbf{x}) \max_{\mathbf{y}} p_Y(\mathbf{y})\right)^{1/n} = \frac{1}{\sqrt[n]{|S_X| \cdot |S_Y|}},$$

and

$$\frac{1}{n} \sum_{\mathbf{x},\mathbf{y} \in \{0,1\}^n} p_X(\mathbf{x}) p_Y(\mathbf{y}) \mathbf{x}^{\mathrm{T}} \mathbf{y} = \frac{\sum_{\mathbf{x} \in S_X, \mathbf{y} \in S_Y} \mathbf{x}^{\mathrm{T}} \mathbf{y}}{n|S_X| \cdot |S_Y|}.$$

Define a new problem as follows:

Problem 2. *For any given $c \in (0, 1/4]$ and every positive integer n, consider the following programming*

$$\begin{aligned} \min_{S_X, S_Y} \quad & \frac{1}{\sqrt[n]{|S_X| \cdot |S_Y|}}, \\ \text{s.t.} \quad & \frac{1}{n|S_X| \cdot |S_Y|} \sum_{\mathbf{x} \in S_X, \mathbf{y} \in S_Y} \mathbf{x}^{\mathrm{T}} \mathbf{y} \le c, \end{aligned} \tag{4}$$

where S_X and S_Y are subsets of $\{0,1\}^n$. Let P'_n be the optimal value of the above program. We are interested in the limit of the sequence $\{P'_n\}$ when $n \to \infty$.

It is obvious that $P_n \le P'_n$ since only distributions that are uniform over support are considered in Problem 2. The following theorem enables us to focus on $\lim_n P'_n$.

Theorem 2. $\lim_{n\to\infty} P'_n / P_n = 1$.

B. Profiles

To study the properties of a set of binary vectors, we introduce the concept of *profile*. For any positive integer m, we call vector $a = (a_1, a_2, \cdots, a_m) \in [0,1]^m$ a *profile* or an *m-profile*. For each $S \subseteq \{0,1\}^n$, define the *profile* of set S as

$$\Gamma(S) = \begin{cases} \dfrac{1}{|S|} \displaystyle\sum_{s \in S} s, & |S| > 0; \\ (0, 0, \ldots, 0), & |S| = 0. \end{cases}$$

We see that $\Gamma(S)$ is an n-profile.

Define the *characteristic function* of an m-profile a as $f_a : [0,1] \to [0,1]$ such that

$$f_a(t) = \begin{cases} a_1, & t = 0; \\ a_{\lceil tm \rceil}, & \forall 0 < t \le 1. \end{cases}$$

The characteristic function of a profile is a step function. For two profiles a and b, we say $a \le b$ if for any $0 \le r \le 1$, $f_a(r) \le f_b(r)$, where a and b may not include the same number of components. For a vector a, we denote by a_i the i-th component of a.

Lemma 1. *For two n profiles a and b, $\frac{1}{n} a^{\mathrm{T}} b = \int_0^1 f_a(t) f_b(t) \mathrm{d}t$.*

Proof: We write according to the definition that

$$\begin{aligned} \frac{1}{n} a^{\mathrm{T}} b &= \frac{1}{n} \sum_{i=1}^{n} a_i b_i \\ &= \frac{1}{n} \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} f_a(t) f_b(t) \mathrm{d}t \\ &= \int_0^1 f_a(t) f_b(t) \mathrm{d}t, \end{aligned}$$

where the second equality holds due to the fact that the characteristic function of a profile is a step function. ∎

The following lemma tells us how to represent the constraint in Problem 2 in a simple way using profiles.

Lemma 2. *In Problem 2, the left hand side of constraint (4) can be expressed as*

$$\frac{1}{n|S_X| \cdot |S_Y|} \sum_{\mathbf{x} \in S_X, \mathbf{y} \in S_Y} \mathbf{x}^{\mathrm{T}} \mathbf{y} = \frac{1}{n} a^{\mathrm{T}} b,$$

where $a = \Gamma(S_X)$ and $b = \Gamma(S_Y)$.

Proof: We can write

$$\begin{aligned} &\frac{1}{n} \cdot \frac{1}{|S_X|} \cdot \frac{1}{|S_Y|} \sum_{\mathbf{x} \in S_X, \mathbf{y} \in S_Y} \mathbf{x}^{\mathrm{T}} \mathbf{y} \\ &= \frac{1}{n} \cdot \frac{1}{|S_X|} \cdot \frac{1}{|S_Y|} \left(\sum_{\mathbf{x} \in S_X} \mathbf{x} \right)^{\mathrm{T}} \left(\sum_{\mathbf{y} \in S_Y} \mathbf{y} \right) \\ &= \frac{1}{n} \cdot \frac{1}{|S_X|} \cdot \frac{1}{|S_Y|} (|S_X| a)^{\mathrm{T}} (|S_Y| b) \qquad (5) \\ &= \frac{1}{n} a^{\mathrm{T}} b, \end{aligned}$$

where (5) follows from the definition of the profile of a set of binary vectors. ∎

The following theorem states that to get the value of P'_n, we only need to consider S_X and S_Y with certain monotone property of their profiles.

Theorem 3. *For all n, there exist $S_X, S_Y \subseteq \{0,1\}^n$ that achieve P'_n in Problem 2 such that for $a = \Gamma(S_X)$ and $b = \Gamma(S_Y)$, $0.5 \ge a_1 \ge a_2 \ge \ldots \ge a_n \ge 0$ and $0 \le b_1 \le b_2 \le \ldots \le b_n \le 0.5$.*

By Theorem 3, it is sufficient for us to consider only profiles $a \in [0, 0.5]^m$. For each m-profile a, define its *n-volume* to be

$$V_n(a) = \max \left\{ |S| : S \subseteq \{0,1\}^n,\ \Gamma(S) \le a \right\}, \tag{6}$$

where n may not be the same as m.

Lemma 3. *For any two profiles p and q, if $p \le q$, we have $V_n(p) \le V_n(q)$ for every positive integer n.*

Proof: Notice that for any n, any *n-profile* smaller than p is smaller than q, then the lemma suffices. ∎

The following theorem gives an upper bound on the volume of a profile, which will be used in the proof of the lower bound on P'_n.

Theorem 4. *Fix an integer m and let $a \in [0, 0.5]^m$ be an m-profile. For any positive integer n, the n-volume of profile a satisfies*

$$V_n(a) \le 2^{\frac{n}{m}\left(\sum_{i=1}^{m} h_{\mathrm{b}}(a_i) + o(1)\right)}, \tag{7}$$

where h_{b} is the binary entropy function defined in (3) *and $o(1) \to 0$ as $n \to \infty$.*

C. Converse and Achievability

Theorem 5. *For any sequence of $S_X, S_Y \subseteq \{0,1\}^n$ such that*

$$\frac{1}{n|S_X|\cdot|S_Y|}\sum_{\mathbf{x}\in S_X,\mathbf{y}\in S_Y}\mathbf{x}^{\mathrm{T}}\mathbf{y}\le c_Q,$$

we have

$$\liminf_{n\to\infty}\frac{1}{\sqrt[n]{|S_X||S_Y|}}\ge 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}.$$

We then give a construction of S_X and S_Y to show that the bound in Theorem 5 is tight.

Theorem 6. *There exists a sequence of $S_X, S_Y \subseteq \{0,1\}^n$ such that*

$$\frac{1}{n|S_X|\cdot|S_Y|}\sum_{\mathbf{x}\in S_X,\mathbf{y}\in S_Y}\mathbf{x}^{\mathrm{T}}\mathbf{y}\le c_Q,$$

and

$$\lim_{n\to\infty}\frac{1}{\sqrt[n]{|S_X||S_Y|}}=4^{-h_{\mathrm{b}}(\sqrt{c_Q})}.$$

Now we are ready to prove Theorem 1.

Proof of Theorem 1: Theorem 5 implies that

$$\liminf_{n\to\infty}P'_n\ge 4^{-h_{\mathrm{b}}(\sqrt{c_Q})},$$

and Theorem 6 implies that

$$\limsup_{n\to\infty}P'_n\le 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}.$$

Thus $\lim_{n\to\infty}P'_n=4^{-h_{\mathrm{b}}(\sqrt{c_Q})}$, which together with Theorem 2 proves Theorem 1. ∎

III. Proofs

A. Proof of Proposition 1

The lower bound follows from $\max_{\mathbf{x}}p_X(\mathbf{x})\ge 1/2^n$ for any distribution p_X over $\{0,1\}^n$. To prove the upper bound, consider the following two distributions:

$$p_X(\mathbf{x})=\begin{cases}1-2c, & \mathbf{x}=\mathbf{0}\\ 2c/(2^n-1), & \mathbf{x}\neq\mathbf{0},\end{cases}$$

where $c\in(0,1/4]$ as given in Problem 2, and $p_Y(\mathbf{y})=1/2^n$ for all $\mathbf{y}\in\{0,1\}^n$. We the have

$$\begin{aligned}&\frac{1}{n}\sum_{\mathbf{x},\mathbf{y}\in\{0,1\}^n}p_X(\mathbf{x})p_Y(\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y}\\&=\frac{2c}{2^n(2^n-1)}\cdot\frac{1}{n}\sum_{\mathbf{x},\mathbf{y}\in\{0,1\}^n}\mathbf{x}^{\mathrm{T}}\mathbf{y}\\&=\frac{c}{2^{n-1}(2^n-1)}2^{2(n-1)}\le c,\end{aligned}$$

and

$$\begin{aligned}P_n&\le\left(\max_{\mathbf{x}}p_X(\mathbf{x})\max_{\mathbf{y}}p_Y(\mathbf{y})\right)^{1/n}\\&=\frac{1}{2}\left(\max\{1-2c,2c/(2^n-1)\}\right)^{1/n}\\&=\frac{1}{2}(1-2c)^{1/n}<\frac{1}{2},\end{aligned}$$

where the second equality follows from $c\le 1/4$ and the last inequality follows from $c>0$.

B. Proof of Theorem 2

Suppose that p_X and p_Y on $\{0,1\}^n$ achieve the minimum objective value P_n in Problem 1. Write

$$\sum_{\mathbf{x},\mathbf{y}\in\{0,1\}^n}p_X(\mathbf{x})p_Y(\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y}=\sum_{\mathbf{x}}p_X(\mathbf{x})\theta_{p_Y}(\mathbf{x}),$$

where

$$\theta_{p_Y}(\mathbf{x})=\mathbf{x}^{\mathrm{T}}\left(\sum_{\mathbf{y}}p_Y(\mathbf{y})\mathbf{y}\right).$$

Let $P_X=\max_{\mathbf{x}}p_X(\mathbf{x})$. We know that $P_X>0$. If $P_X=1$, then there exists $\mathbf{x}_0$ such that $p_X(\mathbf{x}_0)=1$. In this case, $P_n=1/2$ since otherwise we may instead choose p_X such that $p_X(\mathbf{0})=1$ and p_Y such that $p_Y(\mathbf{y})=1/2^n$ for all $\mathbf{y}\in\{0,1\}^n$. Thus we have a contradiction to $P_n<1/2$ (see Proposition 1). Therefore, $0<P_X<1$.

Now consider the following linear program:

$$\begin{aligned}\min_{p_X}\quad&\sum_{\mathbf{x}}p_X(\mathbf{x})\theta_{p_Y}(\mathbf{x}),\\ \text{s.t.}\quad&p_X(\mathbf{x})\le P_X,\ \forall\mathbf{x}\in\{0,1\}^n.\end{aligned}\tag{8}$$

Let p_X^* be an optimal distribution that minimizes the objective of (8). Since the linear program must achieve its optimal value at the extreme points, there must be $\lfloor\frac{1}{P_X}\rfloor$ sequences $\mathbf{x}$ with $p_X^*(\mathbf{x})=P_X$ and one sequence $\mathbf{z}$ with $p_X^*(\mathbf{z})=1-\lfloor\frac{1}{P_X}\rfloor P_X$. For any other sequence $\mathbf{x}$, we have $p_X^*(\mathbf{x})=0$.

We then have

$$\sum_{\mathbf{x}}p_X^*(\mathbf{x})\theta_{p_Y}(\mathbf{x})\le\sum_{\mathbf{x}}p_X(\mathbf{x})\theta_{p_Y}(\mathbf{x})\le nc,$$

and

$$\left(\max_{\mathbf{x}}p_X^*(\mathbf{x})\max_{\mathbf{y}}p_Y(\mathbf{y})\right)^{1/n}=\left(P_X\cdot\max_{\mathbf{y}}p_Y(\mathbf{y})\right)^{1/n}=P_n.$$

Therefore, p_X^* and p_Y also obtain the minimum objective value P_n in Problem 1.

Let S_X be the support of p_X^*. We have $|S_X|=\lceil\frac{1}{P_X}\rceil$, and for any $\mathbf{x}\in S_X$, $\theta_{p_Y}(\mathbf{z})\ge\theta_{p_Y}(\mathbf{x})$. Let $\bar{p}_X$ be the uniform distribution over $S_X\backslash\{\mathbf{z}\}$. Notice for all $\mathbf{x}\in S_X\backslash\{\mathbf{z}\}$,

$$\bar{p}_X(\mathbf{x})\ge p_X^*(\mathbf{x}),$$

and

$$\sum_{\mathbf{x}\in S_X\backslash\{\mathbf{z}\}}(\bar{p}_X(\mathbf{x})-p_X^*(\mathbf{x}))=p_X^*(\mathbf{z}).$$

We have

$$\begin{aligned}&\sum_{\mathbf{x},\mathbf{y}}\bar{p}_X(\mathbf{x})p_Y(\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y}-\sum_{\mathbf{x},\mathbf{y}}p_X^*(\mathbf{x})p_Y(\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y}\\=&\sum_{\mathbf{x}\in S_X\backslash\{\mathbf{z}\}}(\bar{p}_X(\mathbf{x})-p_X^*(\mathbf{x}))\theta_{p_Y}(\mathbf{x})-p_X^*(\mathbf{z})\theta_{p_Y}(\mathbf{z})\\\le&\sum_{x\in S_X\backslash\{\mathbf{z}\}}(\bar{p}_X(\mathbf{x})-p_X^*(\mathbf{x}))\theta_{p_Y}(\mathbf{z})-p_X^*(\mathbf{z})\theta_{p_Y}(\mathbf{z})\\=&\ 0.\end{aligned}$$

Thus

$$\sum_{\mathbf{x},\mathbf{y}}\bar{p}_X(\mathbf{x})p_Y(\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y}\le\sum_{\mathbf{x},\mathbf{y}}p_X^*(\mathbf{x})p_Y(\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y}\le nc.\tag{9}$$

Let $P_n^\dagger = \min_{p_X,p_Y} \left(\max_{\mathbf{x}} p_X(\mathbf{x}) \max_{\mathbf{y}} p_Y(\mathbf{y})\right)^{1/n}$ such that p_X and p_Y satisfy the constraint of Problem 1 and p_X is uniform over its support. We have

$$\begin{aligned} P_n \le P_n^\dagger &\le \left(\max_{\mathbf{x}} \bar{p}_X(\mathbf{x}) \max_{\mathbf{y}} p_Y(\mathbf{y})\right)^{1/n} \\ &= \left(\frac{1}{\lfloor 1/P_X \rfloor} \max_{\mathbf{y}} p_Y(\mathbf{y})\right)^{1/n} \\ &\le \left(\frac{1}{\lceil 1/P_X \rceil - 1} \max_{\mathbf{y}} p_Y(\mathbf{y})\right)^{1/n} \\ &\le \left(3P_X \max_{\mathbf{y}} p_Y(\mathbf{y})\right)^{1/n} \\ &= 3^{1/n} P_n, \end{aligned}$$

where the second inequality follows from $\bar{p}_X$ and p_Y satisfy the constraint of Problem 1 (see (9)), and the last inequality follows from $0 < P_X < 1$ and Lemma 4 (to be proved later in this section). Therefore, $\lim_{n\to\infty} P_n^\dagger / P_n = 1$.

Similar technique can be used to show that $\lim_{n\to\infty} P_n'/P_n^\dagger = 1$, which completes the proof of this theorem. Specifically, suppose that p_X, p_Y on $\{0,1\}^n$ achieve $P_n^\dagger$ where p_X is uniform on its support. Define $P_Y = \max_{\mathbf{y}} p_Y(\mathbf{y})$ and $P_X = \max_{\mathbf{x}} p_X(\mathbf{x})$. Similar to the above argument, there exists distribution p_Y^* such that

1) for $\lfloor \frac{1}{P_Y} \rfloor$ sequences $\mathbf{y}$, $p_Y^*(\mathbf{y}) = P_Y$, for another one sequence $\mathbf{y}_0$, $p_Y^*(\mathbf{y}_0) = 1 - \lfloor \frac{1}{P_Y} \rfloor P_Y$, and for all other sequences $\mathbf{y}$, $p_Y^*(\mathbf{y}) = 0$;
2) $\sum_{\mathbf{x},\mathbf{y}} p_X(\mathbf{x}) p_Y^*(\mathbf{y}) \mathbf{x}^{\mathrm{T}}\mathbf{y} \le \sum_{\mathbf{x},\mathbf{y}} p_X(\mathbf{x}) p_Y(\mathbf{y}) \mathbf{x}^{\mathrm{T}}\mathbf{y} \le nc$; and
3) $(\max_{\mathbf{x}} p_X(\mathbf{x}) \max_{\mathbf{y}} p_Y^*(\mathbf{y}))^{1/n} = (P_X P_Y)^{1/n}$.

Let the support set of distributions p_Y^* be S_Y, and let $\bar{p}_Y$ be the uniform distribution over $S_Y \backslash \{\mathbf{y}_0\}$. Similar to the reasoning of (9), we have

$$\sum_{\mathbf{x},\mathbf{y}} p_X(\mathbf{x}) \bar{p}_Y(\mathbf{y}) \mathbf{x}^{\mathrm{T}}\mathbf{y} \le \sum_{\mathbf{x},\mathbf{y}} p_X(\mathbf{x}) p_Y^*(\mathbf{y}) \mathbf{x}^{\mathrm{T}}\mathbf{y} \le nc.$$

Again, according to Lemma 4,

$$\begin{aligned} P_n^\dagger \le P_n' &\le \left(P_X \max_{\mathbf{y}} \bar{p}_Y(\mathbf{y})\right)^{1/n} \\ &= \left(P_X \frac{1}{\lfloor 1/P_Y \rfloor}\right)^{1/n} \\ &\le \left(P_X \frac{1}{\lceil 1/P_Y \rceil - 1}\right)^{1/n} \\ &\le (3P_X P_Y)^{1/n} \\ &= \sqrt[n]{3} P_n^\dagger, \end{aligned}$$

and hence $\lim_{n\to\infty} P_n'/P_n^\dagger = 1$.

Lemma 4. *For every $x \in (0,1)$,*

$$x(\lceil 1/x \rceil - 1) \ge \frac{1}{3}.$$

Proof: If $x \ge \frac{1}{3}$, then

$$x(\lceil 1/x \rceil - 1) \ge x \ge \frac{1}{3}.$$

If $x < \frac{1}{3}$, then

$$x(\lceil 1/x \rceil - 1) \ge x(1/x - 2) \ge 1 - 2x > \frac{1}{3}.$$

■

C. Proof of Theorem 3

We first show that we only need to consider S_X and S_Y with profiles $a, b \in [0, 0.5]^n$. Suppose that for some i we have $a_i > \frac{1}{2}$. We obtain a new set S_X' by flipping the i-th bit of all vectors in S_X. Let $a' = \Gamma(S_X')$. We have $a_k' = a_k$ for $k \ne i$ and $a_i' = 1 - a_i$. We know from Lemma 2 that for the constraint (4) still holds with S_X' in place of S_X since $a_i' < 0.5 < a_i$. While the objective function of Problem 2 with S_X' in place of S_X does not change since $|S_X'| = |S_X|$. Similarly we can modify S_Y such that all $b_i \le \frac{1}{2}$.

Without the loss of generality, we assume $a_1 \ge a_2 \ge \cdots \ge a_n$. Otherwise we just change the order of the bit in the string. Now we put $b_1, ..., b_n$ in a non-decreasing reordering as: $b_1' \le ... \le b_n'$. There must exist set $S_Y' \subseteq \{0,1\}^n$ such that $\Gamma(S_Y') = (b_1', ..., b_n')^{\mathrm{T}}$ by changing the order of the bits for each string in set S_Y. Then we have

$$\frac{1}{n|S_X||S_Y'|} \sum_{x\in S_X, y \in S_Y'} x^{\mathrm{T}} y = \sum_{i=1}^n a_i b_i' \le \sum_{i=1}^n a_i b_i \le c. \quad (10)$$

The proof is completed by $|S_X||S_Y'| = |S_X||S_Y|$.

D. Proof of Theorem 4

The logarithm in this proof has base 2. Consider subset $S \subset \{0,1\}^n$ with $\Gamma(S) \le a$. Define a random vector $X = (X_1, X_2, \ldots, X_n)$ over $\{0,1\}^n$ with support S and $\Pr\{X = \mathbf{x}\} = \frac{1}{|S|}$ for each $\mathbf{x} \in S$. Recall that the i-th component of $\mathbf{x} \in \{0,1\}^n$ is denoted by $\mathbf{x}_i$. Let $l_k = \lfloor \frac{kn}{m} \rfloor$ for $k = 0, 1, \ldots, m$. Since $(\mathrm{E}[X_1], \mathrm{E}[X_i], \ldots, \mathrm{E}[X_n]) = \Gamma(S) \le a$, we have for $k = 1, \ldots, m$ and $i = 1, \ldots, l_k - l_{k-1}$, $\mathrm{E}[X_{l_{k-1}+i}] = f_{\Gamma(S)}(\frac{l_{k-1}+i}{n}) \le f_a(\frac{l_{k-1}+i}{n}) = a_k$. Note that X_i is a binary random variable. Hence the entropy $H(X_{l_{k-1}+i}) \le h_{\mathrm{b}}(a_k)$ for $k = 1, \ldots, m$ and $i = 1, \ldots, l_k - l_{k-1}$. Therefore,

$$\begin{aligned} \log|S| = H(X) &\le \sum_{k=1}^m \sum_{i=1}^{l_k - l_{k-1}} H(X_{l_{k-1}+i}) \\ &\le \sum_{k=1}^m (l_k - l_{k-1}) h_{\mathrm{b}}(a_k) \\ &\le \frac{n}{m}\left(\sum_i h_{\mathrm{b}}(a_i) + o(1)\right), \end{aligned}$$

where the last inequality follows from $l_k - l_{k-1} \le \frac{n}{m} + 1$ and $o(1)$ tends to zero as n tends to ∞. Since the above inequality holds for all subset $S \subset \{0,1\}^n$ with $\Gamma(S) \le a$, we have

$$V_n(a) \le 2^{\frac{n}{m}(\sum_i h_{\mathrm{b}}(a_i) + o(1))}.$$

E. Proof of Theorem 5

Let $a = \Gamma(S_X)$, $b = \Gamma(S_Y)$. By Theorem 3, it is sufficient for us to consider S_X and S_Y such that $0.5 \geq a_1 \geq \ldots \geq a_n \geq 0$ and $0 \leq b_1 \leq \ldots \leq b_n \leq 0.5$. Hence f_a is decreasing on $[0,1]$, and f_b is increasing on $[0,1]$.

Define two *m-profile*s $\bar{a}$ and $\underline{a}$ such that for $1 \leq i \leq m$,

$$\bar{a}_i = \frac{\lceil m f_a\left(\frac{i-1}{m}\right)\rceil}{m},\ \underline{a}_i = \frac{\lfloor m f_a\left(\frac{i}{m}\right)\rfloor}{m}.$$

We have $f_{\bar{a}}$ and $f_{\underline{a}}$ are decreasing on $[0,1]$.

Lemma 5. *$\underline{a} \leq a \leq \bar{a}$.*

Proof: Notice that f_a is a decreasing function. For every $0 \leq r \leq 1$,

$$f_{\bar{a}}(r) = \bar{a}_{\lceil rm\rceil} \geq f_a\left(\frac{\lceil rm\rceil - 1}{m}\right) \geq f_a(r),$$

and similarly,

$$f_{\underline{a}}(r) = \underline{a}_{\lceil rm\rceil} \leq f_a\left(\frac{\lceil rm\rceil}{m}\right) \leq f_a(r).$$

Thus $\underline{a} \leq a \leq \bar{a}$. ∎

Define two *m-profile*s $\bar{b}$ and $\underline{b}$ such that for $1 \leq i \leq m$,

$$\bar{b}_i = \frac{\lceil m f_b\left(\frac{i}{m}\right)\rceil}{m},\ \underline{b}_i = \frac{\lfloor m f_b\left(\frac{i-1}{m}\right)\rfloor}{m}.$$

We have $f_{\bar{b}}$ and $f_{\underline{b}}$ are increasing on $[0,1]$, and similar to Lemma 5, we have the following lemma.

Lemma 6. *$\underline{b} \leq b \leq \bar{b}$.*

Now we can prove the following lemma.

Lemma 7. *For $m \geq 2$,*

$$\frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i - \frac{1}{n}\sum_{i=1}^{n} a_i b_i < \frac{2}{m}. \tag{11}$$

Proof: Observe that

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i - \frac{1}{n}\sum_{i=1}^{n} a_i b_i &= \frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i - \int_0^1 f_a(t) f_b(t)\mathrm{d}t \\ &\leq \frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i - \int_0^1 f_{\underline{a}}(t) f_{\underline{b}}(t)\mathrm{d}t \\ &= \frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i - \frac{1}{m}\sum_{i=1}^{m} \underline{a}_i\underline{b}_i,\end{aligned}$$

where in the first and last equality we apply Lemma 1. The first equality comes from the fact that f_a and f_b are step functions. By definition, we have for $1 \leq i \leq m-1$, $m\underline{a}_i \geq m\bar{a}_{i+1} - 1$ and $m\underline{b}_{i+1} \geq m\bar{b}_i - 1$. Hence

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i - \frac{1}{m}\sum_{i=1}^{m} \underline{a}_i\underline{b}_i \\ &\leq \frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i - \frac{1}{m}\sum_{i=2}^{m-1}\left(\bar{a}_{i+1} - \frac{1}{m}\right)\left(\bar{b}_{i-1} - \frac{1}{m}\right) \\ &= \frac{1}{m}\Bigg(\bar{a}_1\bar{b}_1 + \bar{a}_2\bar{b}_2 + \sum_{i=3}^{m}\bar{a}_i(\bar{b}_i - \bar{b}_{i-2}) + \\ &\qquad \sum_{i=2}^{m-1}\left(\frac{\bar{a}_{i+1}}{m} + \frac{\bar{b}_{i-1}}{m} - \frac{1}{m^2}\right)\Bigg) \\ &\leq \frac{1}{m}\Bigg(0.25 + 0.25 + \sum_{i=3}^{m} 0.5(\bar{b}_i - \bar{b}_{i-2}) + \\ &\qquad \sum_{i=2}^{m-1}\left(\frac{0.5}{m} + \frac{0.5}{m}\right)\Bigg) \\ &= \frac{1}{m}\left(1.5 - \frac{2}{m} + 0.5\bar{b}_m + 0.5\bar{b}_{m-1} - 0.5\bar{b}_2 - 0.5\bar{b}_1\right) \\ &\leq \frac{2}{m},\end{aligned}$$

where we use the fact that $\bar{a}_i, \bar{b}_i \leq 0.5$. ∎

By Lemma 7 and the condition of the theorem (using the form given in Lemma 2), we have

$$\frac{1}{m}\sum_{i=1}^{m} \bar{a}_i\bar{b}_i \leq \frac{1}{n}\sum_{i=1}^{n} a_i b_i + \frac{2}{m} \leq c_Q + \frac{2}{m}. \tag{12}$$

From Lemma 3 and Theorem 4, we know that

$$\begin{aligned}|S_X||S_Y| &= V_n(a)V_n(b) \\ &\leq V_n(\bar{a})V_n(\bar{b}) \\ &\leq 2^{\frac{n}{m}\left(\sum_{i=1}^{m}\left(h_{\mathrm{b}}(\bar{a}_i) + h_{\mathrm{b}}(\bar{b}_i)\right) + o(1)\right)},\end{aligned}$$

where $o(1) \to 0$ as $n \to \infty$. For $0 \leq t \leq 0.25$, define

$$f(t) = \max_{2t \leq x \leq \frac{1}{2}}\left(h_{\mathrm{b}}(x) + h_{\mathrm{b}}\left(\frac{t}{x}\right)\right). \tag{13}$$

Some properties of the above function are given in Appendix II (see Lemma 8 – 10). We have

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}\left(h_{\mathrm{b}}(\bar{a}_i) + h_{\mathrm{b}}(\bar{b}_i)\right) &= \frac{1}{m}\sum_{i=1}^{m}\left(h_{\mathrm{b}}(\bar{a}_i) + h_{\mathrm{b}}\left(\frac{\bar{a}_i\bar{b}_i}{\bar{b}_i}\right)\right) \\ &\leq \frac{1}{m}\sum_{i=1}^{m} f(\bar{a}_i\bar{b}_i) \leq f\left(c_Q + \frac{2}{m}\right),\end{aligned}$$

where the first inequality follows from the definition of $f(\bar{a}_i\bar{b}_i)$ and the second inequality is obtained by applying (12) and Lemma 10.

Thus for any sufficiently large m,

$$\liminf_{n\to\infty} \frac{1}{\sqrt[n]{|S_X||S_Y|}} \geq 2^{-f\left(c_Q + \frac{2}{m}\right)}.$$

Take $m \to \infty$ we have

$$\liminf_{n\to\infty} \frac{1}{\sqrt[n]{|S_X||S_Y|}} \geq 2^{-f(c_Q)} = 4^{-h_{\mathrm{b}}(\sqrt{c_Q})}, \tag{14}$$

where the last equality is implied by Lemma 8.

F. Proof of Theorem 6

For every n, let

$$S_X = S_Y = \{\mathbf{x} \in \{0,1\}^n : \mathbf{x} \text{ includes at most } n\sqrt{c_Q}\ 1s\}.$$

Then

$$|S_X| = |S_Y| = \sum_{i=0}^{\lfloor n\sqrt{c_Q}\rfloor} \binom{n}{i} = 2^{n(h_{\rm b}(\sqrt{c_Q})+o(1))}, \quad (15)$$

where $o(1) \to 0$ as $n \to \infty$. Thus

$$\lim_{n\to\infty} \frac{1}{\sqrt[n]{|S_X||S_Y|}} = \frac{1}{2^{2h_{\rm b}(\sqrt{c_Q})}} = 4^{-h_{\rm b}(\sqrt{c_Q})}.$$

From the constructions of S_X and S_Y, we know that

$$\Gamma(S_X) = \Gamma(S_Y) \le \left(\sqrt{c_Q}, \sqrt{c_Q}, \cdots, \sqrt{c_Q}\right).$$

Therefore

$$\begin{aligned}\frac{1}{n}\sum_{\mathbf{x}\in S_X, \mathbf{y}\in S_Y} \frac{1}{|S_X||S_Y|}\mathbf{x}^{\rm T}\mathbf{y} &= \frac{1}{n}(\Gamma(S_X))^{\rm T}\Gamma(S_Y) \\ &\le \frac{1}{n}\sum_{i=1}^{n}(\sqrt{c_Q})^2 \\ &= c_Q.\end{aligned}$$

Thus S_X and S_Y satisfies constraints in Theorem 2.

IV. Concluding Remarks

In this paper, we determine for Problem 1 that when $c = c_Q$

$$\lim_{n\to\infty} P_n = 4^{-h_{\rm b}(\sqrt{c})}, \quad (16)$$

which is of particular interest for quantum information. Note that our technique also shows that (16) holds for $c_Q \le c < 1/4$. However, the existing technique in this paper does not imply (16) for $c < c_Q$, which holds if we can show that $f(t)$ (defined in (13)) is concave in $[0, 0.25]$. But we can only show the concavity of $f(t)$ for the range $[0.0625, 0.25]$ (see Appendix II). Whether $f(t)$ is concave in $[0, 0.25]$ is of certain mathematical interest.

Acknowledgments

We thank Xiongfeng Ma, Xiao Yuan and Zhu Cao for introducing us this problem and providing insightful comments to our work.

Appendix I
Background of the Optimization Problem

A. CHSH Inequality

A Bell test experiment has two spatially separated parties, Alice and Bob, who can randomly choose their devices settings X and Y from set $\{0,1\}$ and generate random output bits A and B, respectively. The Clauser-Horne-Shimony-Holt (CHSH) inequality is that

$$S^{(1)} := \sum_{a,b,x,y\in\{0,1\}} (-1)^{a\oplus b+xy} q_{AB|XY}(a,b|x,y) \le 2, \quad (17)$$

where $\oplus$ denotes the exclusive-or of two bits, and $q_{AB|XY}(a,b|x,y)$ is the probability that outputs a and b are generated when the device settings are x and y. To simplify the notations, we may also write $q_{AB|XY}(a,b|x,y)$ as $q(a,b|x,y)$, and use the similar convention for other probability distributions. The theory of quantum mechanics predicts a maximum value for S of $S_Q = 2\sqrt{2}$.

In a local hidden variable model (LHVM), assume that an adversary Eve controls a variable λ taking discrete values so that

$$q(a,b|x,y) = \sum_{\lambda} q(a|x,\lambda)q(b|y,\lambda)q(\lambda|x,y),$$

where $q(a|x,\lambda)$ (resp. $q(b|y,\lambda)$) is the probability that a is output when the setting of Alice (resp. Bob) is x (resp. y), and $q(\lambda|x,y)$ is the conditional probability distribution of the variable λ given x and y. *Free will* is assumed in the derivation of the CHSH inequality, i.e.,

$$q(\lambda|x,y) = q(\lambda). \quad (18)$$

With this assumption, the inequality (17) holds for any LHVM.

We consider the case that the device settings may not be chosen freely, i.e., (18) may not hold. By the Bayes' law,

$$q(\lambda|x,y) = \frac{q(x,y|\lambda)q(\lambda)}{q(x,y)} = 4q(x,y|\lambda)q(\lambda),$$

where $q(x,y)$ is assumed to be $1/4$ so that Alice and Bob cannot detect the existence of adversary Eve. In this case,

$$S = \sum_{\lambda} S_\lambda q(\lambda), \quad (19)$$

where

$$S_\lambda = 4\sum_{a,b,x,y\in\{0,1\}} (-1)^{a\oplus b+xy} q(a|x,\lambda)q(b|y,\lambda)q(x,y|\lambda).$$

The adversary can pick probabilities $q(\lambda)$, $q(x,y|\lambda)$, $q(a|x,\lambda)$ and $q(b|y,\lambda)$ to fake the violation of a Bell's inequality.

The following randomness measure are used in literature [16], [18], [19]

$$P = \max_{x,y,\lambda} q(x,y|\lambda).$$

Note that P takes values from $1/4$ to 1. When $P = 1/4$, all the device settings are uniformly picked independent of λ. When $P = 1$, for at least one value of λ, the device settings are deterministic.

We are interested in the minimum value of P such that $S \ge S_Q$ for certain LHVMs in the independent device setting scenario, i.e., $q(x,y|\lambda) = q(x|\lambda)q(y|\lambda)$. In other words, we want to solve the following problem

$$\begin{array}{ll}\min & \max_{x,y,\lambda} q(x,y|\lambda) \\ \text{s.t.} & \sum_\lambda S_\lambda q(\lambda) \ge S_Q, \\ & \sum_\lambda q(x,y|\lambda)q(\lambda) = \frac{1}{4}, \\ & q(x,y|\lambda) = q(x|\lambda)q(y|\lambda),\end{array} \quad (20)$$

where the minimization is over all the possible (conditional) distributions $q(\lambda)$, $q(a|x,\lambda)$, $q(b|y,\lambda)$ and $q(x,y|\lambda)$ with $q(x,y|\lambda) = q(x|\lambda)q(y|\lambda)$. Due to the convexity of the constraints with respect to $q(a|x,\lambda)$ and $q(b|y,\lambda)$, we can consider

only deterministic distributions $q(a|x,\lambda)$ and $q(b|y,\lambda)$ without changing the optimal value of (20). Let $a = a(x,\lambda)$ and $b = b(y,\lambda)$. Rewrite

$$S_\lambda = 4 \sum_{x,y\in\{0,1\}} (-1)^{a(x,\lambda)\oplus b(y,\lambda)+xy} q(x,y|\lambda). \tag{21}$$

In the above formulations, only a *single run* of the test is performed. It is more realistic to consider that the device settings in different runs are correlated, which is referred to as the *multiple-run* scenario, where the device settings $\mathbf{x} = (x_1,\ldots,x_n)^{\mathrm{T}}$ and $\mathbf{y} = (y_1,\ldots,y_n)^{\mathrm{T}}$ in n runs of the tests follow a joint distribution $q(\mathbf{x},\mathbf{y}|\lambda)$. Similar to the discussion of the single-run scenario, for multiple runs, we have the CHSH inequality $S^{(n)} = \sum_\lambda S_\lambda^{(n)} q(\lambda) \le 2$ with

$$\begin{aligned} S_\lambda^{(n)} &= \frac{4}{n} \sum_{\mathbf{x},\mathbf{y}\in\{0,1\}^n} q(\mathbf{x},\mathbf{y}|\lambda) \sum_{i=1}^{n} (-1)^{a(x_i,\lambda)\oplus b(y_i,\lambda)+x_i y_i} \\ &= 4 \sum_{\mathbf{x},\mathbf{y}\in\{0,1\}^n} q(\mathbf{x},\mathbf{y}|\lambda) \Big[\pi(0,0|\mathbf{x},\mathbf{y})(-1)^{a(0,\lambda)\oplus b(0,\lambda)} \\ &\quad + \pi(0,1|\mathbf{x},\mathbf{y})(-1)^{a(0,\lambda)\oplus b(1,\lambda)} \\ &\quad + \pi(1,0|\mathbf{x},\mathbf{y})(-1)^{a(1,\lambda)\oplus b(0,\lambda)} \\ &\quad + \pi(1,1|\mathbf{x},\mathbf{y})(-1)^{a(1,\lambda)\oplus b(1,\lambda)+1} \Big] \\ &= 4 \sum_{x,y\in\{0,1\}} (-1)^{a(x,\lambda)\oplus b(y,\lambda)+xy} \pi(x,y|\lambda) \end{aligned} \tag{22}$$

where $\pi(x,y|\mathbf{x},\mathbf{y})$ is the fraction of (x,y) pairs among the pairs $(x_k,y_k), k = 1,\ldots,n$, and

$$\pi(x,y|\lambda) = \sum_{\mathbf{x},\mathbf{y}\in\{0,1\}^n} q(\mathbf{x},\mathbf{y}|\lambda)\pi(x,y|\mathbf{x},\mathbf{y}).$$

Note that (22) shares the same form as (21).

Define the measure of measurement dependence for multiple runs as

$$P^{(n)} = \left(\max_{\mathbf{x},\mathbf{y},\lambda} q(\mathbf{x},\mathbf{y}|\lambda) \right)^{1/n}.$$

Under the independent device setting condition that $q(\mathbf{x},\mathbf{y}|\lambda) = q(\mathbf{x}|\lambda)q(\mathbf{y}|\lambda)$, the problem of interest now becomes

$$\begin{aligned} \min \quad & \left(\max_{\mathbf{x},\mathbf{y},\lambda} q(\mathbf{x},\mathbf{y}|\lambda) \right)^{1/n} \\ \text{s.t.} \quad & \sum_\lambda S_\lambda^{(n)} q(\lambda) \ge S_Q \\ & \textstyle\sum_\lambda q(\mathbf{x},\mathbf{y}|\lambda) q(\lambda) = \frac{1}{4^n}, \\ & q(\mathbf{x},\mathbf{y}|\lambda) = q(\mathbf{x}|\lambda) q(\mathbf{y}|\lambda), \end{aligned} \tag{23}$$

where $S_\lambda^{(n)}$ is defined in (22). Note that when $n = 1$, (23) becomes (20).

B. Simplification

We use the case $n = 1$ to illustrate how to simplify the above optimization problem.

First, we determine the choice of the output functions $a(x,\lambda)$ and $b(x,\lambda)$ using the approach in [18]. For a give value of λ, there are totally 16 different pairs of the output functions (a,b). Table II lists the eight possible output functions with $a(0,\lambda) = 0$. It is not necessary to consider the other eight possible output functions with $a(0,\lambda) = 1$ since they give the same set of S_λ as listed in the last column in Table II. Since the output functions with index $1,2,3,4$ are better than the output functions with index $5,6,7,8$, respective, we use the former four choices of the output functions.

TABLE III
OUTPUT FUNCTION ASSIGNMENT.

λ	$a(0,\lambda)$	$a(1,\lambda)$	$b(0,\lambda)$	$b(1,\lambda)$
0	0	0	0	0
1	0	0	0	1
2	0	1	0	0
3	0	1	1	0

TABLE IV
ASSIGNMENT OF THE DEVICE SETTING DISTRIBUTIONS.

λ	$q(0,0\|\lambda)$	$q(0,1\|\lambda)$	$q(1,0\|\lambda)$	$q(1,1\|\lambda)$
0	$q^*(0,0)$	$q^*(0,1)$	$q^*(1,0)$	$q^*(1,1)$
1	$q^*(1,0)$	$q^*(1,1)$	$q^*(0,0)$	$q^*(0,1)$
2	$q^*(0,1)$	$q^*(0,0)$	$q^*(1,1)$	$q^*(1,0)$
3	$q^*(1,1)$	$q^*(1,0)$	$q^*(0,1)$	$q^*(0,0)$

With the choices of the output functions as specified above, the constraint $\sum_\lambda q(x,y|\lambda)q(\lambda) = \frac{1}{4}$ is redundant. To show this, we consider a LHVM (denoted by L^*) with a constant λ, and output functions $a^*(x) = b^*(y) = 0$. (Other choices of $a^*(x)$ and $b^*(y)$ can be shown similarly.) We use $q^*(x,y)$ to denote the device setting distribution related to this LHVM. Define a new LHVM (denoted by L) with $\lambda = 0,1,2,3$ and $q(\lambda) = 1/4$ as follows: The output functions are assigned according to Table III, and the device setting distributions are assigned according to Table IV. It can be verified that

$$\begin{aligned} P &= \max_{x,y\in\{0,1\},\lambda=\{0,1,2,3\}} q(x,y|\lambda) \\ &= \max_{x,y\in\{0,1\}} q^*(x,y), \end{aligned}$$

and

$$\begin{aligned} S &= \sum_{\lambda=\{0,1,2,3\}} q(\lambda) 4 \sum_{x,y\in\{0,1\}} (-1)^{a(x,\lambda)\oplus b(y,\lambda)+xy} q(x,y|\lambda) \\ &= q^*(0,0) + q^*(0,1) + q^*(1,0) - q^*(1,1). \end{aligned}$$

Hence, if LHVM L^* achieves the optimal value of (20), so does LHVM L, which has $q(x,y) = 1/4$.

Further, for each of the four pairs of output functions with index $1,2,3,4$ in Table II, the corresponding S_λ involves only one summands with negative coefficient. Since the four probability masses $q(0,0|\lambda)$, $q(0,1|\lambda)$, $q(1,0|\lambda)$ and $q(1,1|\lambda)$ are symmetry, these four pairs of output functions achieve the same optimal value. Here we use $a(x,\lambda) = b(y,\lambda) = 0$ so that

$$\sum_\lambda S_\lambda^{(1)} q(\lambda) = 4 - 8 q_{XY}(1,1).$$

With these simplifications, the above minimization problem

TABLE II
OUTPUT FUNCTION ASSIGNMENT.

	$a(0,\lambda)$	$a(1,\lambda)$	$b(0,\lambda)$	$b(1,\lambda)$	$S_\lambda/4$
1	0	0	0	0	$q(0,0\|\lambda)+q(0,1\|\lambda)+q(1,0\|\lambda)-q(1,1\|\lambda)$
2	0	0	0	1	$q(0,0\|\lambda)-q(0,1\|\lambda)+q(1,0\|\lambda)+q(1,1\|\lambda)$
3	0	1	0	0	$q(0,0\|\lambda)+q(0,1\|\lambda)-q(1,0\|\lambda)+q(1,1\|\lambda)$
4	0	1	1	0	$-q(0,0\|\lambda)+q(0,1\|\lambda)+q(1,0\|\lambda)+q(1,1\|\lambda)$
5	0	0	1	0	$-q(0,0\|\lambda)+q(0,1\|\lambda)-q(1,0\|\lambda)-q(1,1\|\lambda)$
6	0	0	1	1	$-q(0,0\|\lambda)-q(0,1\|\lambda)-q(1,0\|\lambda)+q(1,1\|\lambda)$
7	0	1	0	1	$q(0,0\|\lambda)-q(0,1\|\lambda)-q(1,0\|\lambda)-q(1,1\|\lambda)$
8	0	1	1	1	$-q(0,0\|\lambda)-q(0,1\|\lambda)+q(1,0\|\lambda)-q(1,1\|\lambda)$

becomes

$$\begin{aligned} \min \quad & \max_{x,y,\lambda} q(x,y|\lambda) \\ \text{s.t.} \quad & q_{XY}(1,1) \le \frac{4-S_Q}{8}, \\ & q(x,y|\lambda) = q(x|\lambda)q(y|\lambda). \end{aligned} \tag{24}$$

For any λ and $c \in [0,0.5]$, let $P(c)$ be the minimum value of $\max_{x,y} q(x,y|\lambda)$ such that $q(1,1|\lambda) \le c, q(x,y|\lambda) = q(x|\lambda)q(y|\lambda)$. Note that $P(c)$ does not depend on the choices of λ, and $P(c)$ is a non-increasing function of c. It clear that if we use only a constant λ in (24), the optimal value is $P(\frac{4-S_Q}{8})$. Now we show that it is sufficient to consider a constant λ. Suppose that $q^*(x,y|\lambda)$ achieves the optimal value of (24). Let $c_\lambda = q^*(1,1|\lambda)$. By the first constraint of (24), we have $\sum_\lambda q^*(\lambda)c_\lambda = \frac{4-S_Q}{8}$, which implies the existence of certain λ^* such that $c_{\lambda^*} \le \frac{4-S_Q}{8}$. By the definition of $P(c)$, we have

$$\max_{x,y} q^*(x,y|\lambda) \ge P(c_\lambda),$$

which implies

$$\max_{\lambda,x,y} q^*(x,y|\lambda) \ge \max_\lambda P(c_\lambda) \ge P(c_{\lambda^*}) \ge P((4-S_Q)/8).$$

In other words, using a LHVM with λ taking multiple values cannot achieve smaller optimal value than $P(\frac{4-S_Q}{8})$. Hence, it is sufficient to consider a constant λ, and (24) becomes

$$\begin{aligned} \min \quad & \max_{x,y} q(x)q(y) \\ \text{s.t.} \quad & q_X(1)q_Y(1) \le \frac{4-S_Q}{8} \end{aligned}$$

Similar to the reasoning of the single-run case, we can use a deterministic strategy λ with $a(x,\lambda) = b(y,\lambda) = 0$, and simplify problem (23) to

$$\begin{aligned} \min \quad & \left(\max_{\mathbf{x},\mathbf{y}} q(\mathbf{x},\mathbf{y})\right)^{1/n} \\ \text{s.t.} \quad & \frac{1}{n}\sum_{\mathbf{x},\mathbf{y}\in\{0,1\}^n} q(\mathbf{x},\mathbf{y})\mathbf{x}^{\mathrm{T}}\mathbf{y} \le \frac{4-S_Q}{8}, \\ & q(\mathbf{x},\mathbf{y}) = q(\mathbf{x})q(\mathbf{y}), \end{aligned}$$

which is (2).

APPENDIX II
PROPERTIES OF A FUNCTION

We study some properties of the function $f(t)$ defined in (13). Recall that

$$f(t) = \max_{2t\le x\le\frac{1}{2}} \left(h_{\mathrm{b}}(x) + h_{\mathrm{b}}\left(\frac{t}{x}\right)\right), \quad 0 \le t \le 0.25.$$

The next lemma implies that $f(t) = 2h_{\mathrm{b}}(\sqrt{t})$ for $0.0625 \le t \le 0.25$.

Lemma 8. *For $0.0625 \le t \le 0.25$, $2t \le x \le 0.5$, we have*

$$h_{\mathrm{b}}(x) + h_{\mathrm{b}}\left(\frac{t}{x}\right) \le 2h_{\mathrm{b}}(\sqrt{t}),$$

where the equality holds for $x = \sqrt{t}$. That is $f(t) = 2h_{\mathrm{b}}(\sqrt{t})$ for $t \in [0.0625, 0.25]$.

Proof: Fix t. Let $u(x) = h_{\mathrm{b}}(x) + h_{\mathrm{b}}\left(\frac{t}{x}\right)$. Observe that $u(x) = u\left(\frac{t}{x}\right)$. Thus it suffices to show $u(x) \le 2h_{\mathrm{b}}(\sqrt{t})$ for $2t \le x \le \sqrt{t}$. Taking derivative on u we have

$$u'(x) = -\log x + \log(1-x) + \frac{t}{x^2}\log\left(\frac{t}{x}\right) - \frac{t}{x^2}\log\left(1-\frac{t}{x}\right)$$

Let $v(x) = -x\log x + x\log(1-x)$, we have

$$xu'(x) = v(x) - v\left(\frac{t}{x}\right) \tag{25}$$

From $t \ge \frac{1}{16}$ we have

$$\frac{t}{x} \ge \frac{1}{2} - x \ge \frac{1}{4}. \tag{26}$$

We may verify that v is decreasing on $[0.25, 0.5]$. If $x \ge 0.25$, then $xu'(x) \ge 0$ since $x \le \frac{t}{x}$. Otherwise, we may verify $v(x) \ge v(0.5-x)$ for $x \le 0.25$. Then apply (26) to (25) we have

$$xu'(x) = v(x) - v\left(\frac{t}{x}\right) \ge v(x) - v\left(0.5-x\right) \ge 0 \tag{27}$$

Therefore u is an increasing function on $[2t, \sqrt{t}]$, which implies $u(x) \le 2h_{\mathrm{b}}(\sqrt{t})$. ∎

Lemma 9. *Function $f(t)$ is increasing on $[0, 0.25]$.*

Proof: To show that f is increasing, fix any $0 \le t_1 < t_2 \le 0.25$. We write $f(t_1) = h_{\mathrm{b}}(x_1) + h_{\mathrm{b}}(y_1)$ where x_1 maximizes $h_{\mathrm{b}}(x) + h_{\mathrm{b}}\left(\frac{t_1}{x}\right)$ for $x \in [2t_1, 0.5]$ and $x_1 y_1 = t_1$. We know that $0 \le x_1, y_1 \le 0.5$. Find x_2 and y_2 such that $x_1 \le x_2 \le \frac{1}{2}$, $y_1 \le y_2 \le \frac{1}{2}$ such that $x_2 y_2 = t_2$. Therefore

$$f(t_1) = h_{\mathrm{b}}(x_1) + h_{\mathrm{b}}(y_1) \le h_{\mathrm{b}}(x_2) + h_{\mathrm{b}}(y_2) \le f(t_2).$$

∎

Lemma 10. *For any $c' \ge c_Q = \frac{2-\sqrt{2}}{4} \approx 0.1464$, if k real numbers $t_1, t_2, \cdots, t_k \in [0, 0.25]$ such that $\frac{1}{k}\sum_{i=1}^k t_i \le c'$, we have*

$$\frac{1}{k}\sum_{i=1}^k f(t_i) \le f(c').$$

Proof: Let $f_0(t) = 2h_{\rm b}\left(\sqrt{t}\right)$, $0 \le t \le 0.25$. From Lemma 8 $f(t) = f_0(t)$ for $t \ge 0.0625$. Let f_1 be the tangent line of f_0 on $(0.14, f_0(0.14))$. Notice that $h_{\rm b}(x)$ and $\sqrt{x}$ are both concave on their domains. We see that $f_0(t)$ is also concave on $[0, 0.25]$. Observe that f_0 is concave and increasing on $\left[0, \frac{1}{4}\right]$, we have f_1 is an increasing function, while for every $t \in [0, 0.25]$, $f_0(t) \le f_1(t)$.

Let $g(t)$ be a function defined on $[0, 0.25]$ such that

$$g(t) = \begin{cases} f_1(t) & 0 \le t \le 0.14; \\ f_0(t) & 0.14 < t \le 0.25. \end{cases}$$

Observe that g is linear on $[0, 0.14]$ and concave on $[0.14, 0.25]$, thus g is concave on $[0, 0.25]$. For $0 \le t < 0.0625$,

$$\begin{aligned} f(t) &\le f(0.0625) \\ &= f_0(0.0625)\ (= 1.623) \\ &< g(0)\ (= 1.630) \\ &\le g(t). \end{aligned}$$

For $0.0625 \le t \le 0.25$,

$$f(t) = f_0(t) \le g(t).$$

Thus g is always not smaller than f. Take $t'_1, t'_2, \cdots, t'_k \le 0.25$ such that $t_i \le t'_i$ for all $1 \le i \le k$, while $\frac{1}{k}\sum_{i=1}^{k} t'_i = c'$. Applying Jensen's inequality we have

$$\begin{aligned} \frac{1}{k}\sum_{i=1}^{k} f(t_i) &\le \frac{1}{k}\sum_{i=1}^{k} f(t'_i) \\ &\le \frac{1}{k}\sum_{i=1}^{k} g(t'_i) \\ &\le g\left(\frac{1}{k}\sum_{i=1}^{k} t'_i\right) \\ &= g(c') \\ &= f(c'), \end{aligned}$$

where the first inequality holds since f is increasing, the second inequality holds since g is always no less than f, and the last equality follows from $c' \ge c_Q > 0.14$. ∎

REFERENCES

[1] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, vol. 1, no. 3, pp. 195–200, 1964.

[2] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.*, vol. 23, pp. 880–884, Oct 1969.

[3] A. Aspect, P. Grangier, and G. Roger, "Experimental realization of Einstein-Podolsky-Rosen-Bohm *Gedankenexperiment* : A new violation of Bell's inequalities," *Phys. Rev. Lett.*, vol. 49, pp. 91–94, Jul 1982.

[4] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin, and A. Zeilinger, "Bell violation using entangled photons without the fair-sampling assumption," *Nature*, vol. 497, no. 7448, pp. 227–230, 05 2013.

[5] M. Feldmann, "New loophole for the Einstein-Podolsky-Rosen paradox," *Found. Phys. Lett.*, vol. 8, no. 1, pp. 41–53, 1995.

[6] J. Kofler, T. Paterek, and i. c. v. Brukner, "Experimenter's freedom in Bell's theorem and quantum cryptography," *Phys. Rev. A*, vol. 73, p. 022104, Feb 2006.

[7] M. J. W. Hall, "Local deterministic model of singlet state correlations based on relaxing measurement independence," *Phys. Rev. Lett.*, vol. 105, p. 250404, Dec 2010.

[8] J. Barrett and N. Gisin, "How much measurement independence is needed to demonstrate nonlocality?" *Phys. Rev. Lett.*, vol. 106, p. 100406, Mar 2011.

[9] M. J. W. Hall, "Relaxed Bell inequalities and kochen-specker theorems," *Phys. Rev. A*, vol. 84, p. 022102, Aug 2011.

[10] D. Mayers and A. Yao, "Quantum cryptography with imperfect apparatus," in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, ser. FOCS '98. Washington, DC, USA: IEEE Computer Society, 1998, pp. 503–509.

[11] A. Acín, N. Gisin, and L. Masanes, "From Bell's theorem to secure quantum key distribution," *Phys. Rev. Lett.*, vol. 97, p. 120405, Sep 2006.

[12] U. Vazirani and T. Vidick, "Fully device-independent quantum key distribution," *Phys. Rev. Lett.*, vol. 113, p. 140501, Sep 2014.

[13] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, "Random numbers certified by Bell's theorem," *Nature*, vol. 464, no. 7291, pp. 1021–1024, 04 2010.

[14] R. Colbeck and R. Renner, "Free randomness can be amplified," *Nat Phys*, vol. 8, no. 6, pp. 450–453, 06 2012.

[15] C. Dhara, G. de la Torre, and A. Acín, "Can observed randomness be certified to be fully intrinsic?" *Phys. Rev. Lett.*, vol. 112, p. 100402, Mar 2014.

[16] D. E. Koh, M. J. W. Hall, Setiawan, J. E. Pope, C. Marletto, A. Kay, V. Scarani, and A. Ekert, "Effects of reduced measurement independence on Bell-based randomness expansion," *Phys. Rev. Lett.*, vol. 109, p. 160404, Oct 2012.

[17] L. P. Thinh, L. Sheridan, and V. Scarani, "Bell tests with min-entropy sources," *Phys. Rev. A*, vol. 87, p. 062121, Jun 2013.

[18] J. E. Pope and A. Kay, "Limited measurement dependence in multiple runs of a Bell test," *Phys. Rev. A*, vol. 88, p. 032110, Sep 2013.

[19] X. Yuan, Z. Cao, and X. Ma, "Randomness requirement on the Clauser-Horne-Shimony-Holt Bell test in the multiple-run scenario," *Phys. Rev. A*, vol. 91, p. 032111, Mar 2015.

[20] G. Pütz, D. Rosset, T. J. Barnea, Y.-C. Liang, and N. Gisin, "Arbitrarily small amount of measurement independence is sufficient to manifest quantum nonlocality," *Phys. Rev. Lett.*, vol. 113, p. 190402, Nov 2014.

[21] J. Gallicchio, A. S. Friedman, and D. I. Kaiser, "Testing Bell's inequality with cosmic photons: Closing the setting-independence loophole," *Phys. Rev. Lett.*, vol. 112, p. 110405, Mar 2014.